IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 November 2023, accepted 21 November 2023, date of publication 24 November 2023,
date of current version 29 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3336862

RESEARCH ARTICLE

Semantic Segmentation of Gastrointestinal Tract in MRI Scans Using PSPNet Model With ResNet34 Feature Encoding Network

NEHA SHARMA[1], SHEIFALI GUPTA[1], ADEL RAJAB[2], MOHAMED A. ELMAGZOUB[3], KHAIRAN RAJAB[2], AND ASADULLAH SHAIKH[4], (Senior Member, IEEE)

[1]Chitkara University Institute of Engineering and Technology, Chitkara University, Chandigarh, Punjab 140401, India
[2]Department of Computer Science, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia
[3]Department of Network and Communication Engineering, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia
[4]Department of Information Systems, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia

Corresponding author: Sheifali Gupta (sheifali.gupta@chitkara.edu.in)

This work was supported by the Deanship of Scientific Research at Najran University under the Research Groups Funding Program under Grant NU/RG/SERC/12/35.

ABSTRACT Gastrointestinal (GI) cancer is the most common cancer in men and women. GI cancers are increasing every year worldwide. In the biomedical industry, Radiation treatment is a frequent choice for treating cancers of the GI tract in which the oncologist focuses the high range of X-ray beams on the tumor while avoiding the healthy organs. Manual segmentation of healthy organs to focus X-ray beams only on the tumor portion is very tedious and time-consuming, which can lead the treatment from a few minutes to hours. Deep learning techniques can concur with this problem by segmentation of healthy organs. This research article proposes a deep learning-based model Pyramid Scene Parsing Network (PSPNet) for segmenting organs such as the stomach, small bowel, and large bowel in the GI tract. The model has been simulated with five feature encoding networks: ResNext 50, Timm_Gernet_S, ResNet 34, EfficientNet B1, and MobileNet V2. These encoders were used for downsampling the feature map in the PSPNet architecture. The implementations have been performed using the UW Madison GI tract dataset, which contains 38,496 MRI scans of cancer patients. The model was evaluated using validation dice, Jaccard, and validation loss. The results reveal that the PSPNet model combined with ResNet 34 as encoder outperforms the other feature encoding networks with validation dice as 0.8842, validation Jaccard as 0.8531, and validation loss as 0.1365. Radiation oncologists can use the proposed model to speed up radiation therapy for cancer treatment.

INDEX TERMS Segmentation, gastrointestinal tract, PSPNet, encoders, MRI scans, radiation therapy, UW madison.

I. INTRODUCTION

Segmentation of gastrointestinal (GI) tract organs is vital in medical imaging and computer-aided diagnosis, offering profound implications for early detection, diagnosis, and treatment of GI diseases. Gastrointestinal disorders, including gastrointestinal (GI) tumors, colorectal cancer, Crohn's disease, and ulcers, pose significant health risks to individuals worldwide.

Globally, gastrointestinal (GI) tumors are the most prevalent forms of cancer [1], [2]. According to the American Cancer Society, there will be around 26,500 new instances of stomach cancer in the United States in 2023 (15,930 in men and 10,570 in women) and approximately 11,130 fatalities from this kind of cancer (6,690 men and 4,440 women) [3]. GI tract cancer is developing cancerous cells in any part of the gastrointestinal system, including the esophagus, stomach,

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

small bowel, large bowel, liver, pancreas, and anus [4]. The most common GI cancers include colorectal, gastric, and pancreatic cancers. Depending on the location, stage, and kind of cancer, GI cancer treatment often includes surgery, radiation treatment, and chemotherapy [2]. Radiation treatment is a frequent choice for treating cancers of the GI tract. This treatment utilizes high-energy radiation to kill cancer cells [5], [6]. Before starting radiation therapy, a healthcare professional will carefully evaluate a person's medical history, overall health, and cancer stage to determine the best treatment plan. The oncologist focuses the X-ray beams on cancer while avoiding the healthy organs. Radiation is often administered daily over several weeks. Segmenting the healthy organs of the GI tract from cancer is necessary. This can be very tedious if performed manually.

As the prevalence of GI-related illnesses continues to rise, the demand for reliable and automated segmentation methods has grown exponentially. The segmentation of GI tract organs presents a unique set of challenges due to the complex anatomical structures, variability in patient anatomy, and the potential presence of pathologies. Therefore, it is essential to develop advanced computational techniques that can accurately and efficiently delineate the boundaries of GI organs, providing healthcare professionals with precise insights into the patient's condition.

Deep learning is a solution to this problem, which can automatically segment healthy organs to speed up the treatment. In medical imaging, deep learning algorithms may segment or identify specific structures or regions of interest within an image [7], [8], [9]. The segmentation method requires training a model on a vast dataset of annotated images, where the GI tract is designated as a distinct region of interest [10]. The model then applies this training to new images to partition the gastrointestinal tract. This paper proposes a Pyramid Scene Parsing Network (PSPNet) to segment the GI tract's healthy organs using the UW Madison GI tract database.

Further, five different pre-trained models were used as the backbone of the PSPNet model. Experimenting with different transfer learning models allows to gain insights into which architectures work well for your particular segmentation problem. This research can lead to a deeper understanding of the relationship between feature extraction and segmentation performance and guide improvements in future models. The PSPNet model also includes a decoder module to refine segmentation findings and enhance the model's overall performance.

The following are the primary contributions of this manuscript:

- A deep learning model, PSPNet, with a feature encoding network, pyramid pooling module, and decoder, has been proposed for segmenting small bowel, large bowel, and stomach in the GI tract to help radiation oncologists speed up cancer treatment. The significance of the feature encoding network lies in its ability to capture and represent the hierarchical, context-rich information from the input image, which is essential for accurate and fine-grained segmentation.
- Different transfer learning models, such as ResNext 50, Timm_Gernet_S, ResNet 34, EfficientNet B1, and MobileNet V2, have been pre-trained on GI tract datasets for feature encoding network and have learned to capture different types of features. A more comprehensive set of features can be potentially extracted using a variety of these models. This diversity in feature representation can be beneficial for capturing a wide range of visual patterns and semantics in the input images, which may lead to better segmentation performance.
- The evaluation of various feature encoding networks in the PSPNet segmentation model has revealed that ResNet-34 stands out as the top-performing choice, demonstrating superior results in metrics such as the Dice coefficient, Jaccard index, and loss. The performance of different feature encoding networks has been evaluated in which ResNet 34 has performed best in terms of dice coefficient, Jaccard, and loss.

The remaining part of the manuscript is arranged as Section II describes the related work of the gastrointestinal tract. Section III shows the proposed methodology utilized for the segmentation. Section IV represents the dataset used for this study. Section V describes the PSPNet design. The results and discussions are presented in Section VI, and the article is wrapped up in Section VII.

II. RELATED WORK

In recent years, a great deal of study has been conducted on gastrointestinal (GI) tract cancer, which has been classified. In 2012, Li B. et al. addressed the issue of automated tumor detection in Wireless Capsule Endoscopy (WCE) images. To describe wireless capsule endoscopy (WCE) images, texture statistics incorporating uniform local binary patterns (LBP) and wavelets were presented. The suggested attributes are insensitive to changes in light and explain the multi-resolution properties of WCE images. Comprehensive testing shows that the recommended computer-aided diagnostic method has a good % tumor detection accuracy of 92.4% in WCE images [11]. Zhou et al. established a global arithmetical approach for repeatedly detecting cysts and determining their ranges in Video capsule endoscopy (VCE) frames in 2014. The suggested system collects statistical data from RGB channels. The statistical data was then loaded into a Support Vector Machines (SVM) to assess the presence and radius of polyps [12]. Wang et al. introduced "Polyp-Alert," a software solution that provides visual feedback during colonoscopy to aid endoscopists in locating polyps. Polyp-Alert detects a polyp along the shape of a polyp—using our prior edge optical characteristics and a classifier. Using 53 video clips of complete operations, the program accurately recognized 97.7% of polyp shots [13]. In 2017, Li et al. introduced a Convolutional Neural Network (CNN) topology for segmenting colorectal polyps. This approach may generate a prediction map with the exact dimensions as the original image of the input network. They evaluated their strategy using the CVC-ClinicDB database [14]. Nguyen et al. suggested an encoder-decoder network-based polyp segmentation approach. They make an accurate forecast by integrating many models to equate the likelihood of each image created using the model. The suggested strategy outperforms state-of-the-art findings, according to an evaluation utilizing the ETIS-LariPolypDB database [15]. In 2019, Dijkstra et al. reported a one-shot approach for characterizing polyps in colonoscopy images. For semantic segmentation, they employ a CNN model. The network was tested on publicly available datasets and yielded encouraging results [16]. Nguyen et al. introduced MED-Net, a novel polyp segmentation approach based on the construction encoder-decoder model, in 2020. Moreover, they provide a supplementary technique for boosting the system's segmentation performance using an augmentation and loss function [17]. Jha et al. published the "Kvasir-Instrument" dataset in 2022, with almost 600 frames featuring GI operation equipment. Two professional GI endoscopists validated the dataset, which contains masks in addition to the images. Furthermore, they serve as a foundation for segmenting GI models to boost research [18]. Sharma et al. employed a conventional U-Net design with a different encoder in 2022. More sophisticated algorithms have shown excellent results in many categorization problems. These algorithms can be used as encoders [19]. In 2022, Ye et al. suggested the SIA-Unet, an upgraded model with Magnetic resonance imaging (MRI) images. SIA-Unet additionally contains an attention tool that filters the spatial information of the feature map to extract relevant data. Comprehensive tests on the UW-Madison dataset were carried out to assess the performance of SIA-Unet [20]. In 2022, Nemani et al. presented a hybrid CNN-transformer architecture for segmenting distinct organs from images. With Dice and Jaccard coefficients of 0.79 and 0.72, the suggested approach proved resilient, scalable, and computationally efficient. The suggested method also illustrates the idea of deep automation to increase treatment efficacy [21]. In 2022, Chou et al. used U-Net and Mask R-CNN approaches to segment different regions. For the validation set, the top Mask R-CNN model received a Dice score of 0.73 [22]. To conduct pixel-level image classification and segmentation, an encoder for the U-Net model, a U-Net decoder, and feature fusion architecture are all components of the network model for GI segmentation that Niu et al. published in 2022. The experimental findings demonstrate that their model improves its Intersection over union (IOU) as matched to other networks [23]. In 2022, Li et al. developed an enhanced 2.5D approach for GI Tract image segmentation. They suggested a technique for combining 2.5D and 3D findings. The findings combination approach enhances scores by 0.007 compared to 2.5D and by 0.009 compared to 3D [24]. In 2022, Chia et al. introduced two baseline methods: a UNet trained on a ResNet50 backbone and a more economical and streamlined UNet. They also examine Feature-wise Linear Modulation (FiLM), a way of improving the UNet model by adding image metadata such as the position of the MRI scan cross-section and the pixel height and breadth [25]. Georgescu et al. suggested a unique technique for generating ensembles of diverse architectures for medical image segmentation that uses the diversity of the models in the ensemble. They run gastrointestinal tract image segmentation studies to compare their diversity-promoting ensemble (DiPE) with another technique for creating ensembles that rely on picking the highest-scoring U-Net models. Their empirical data demonstrate that DiPE outperforms individual models and the ensemble building technique based on selecting the highest-scoring models [26].

III. PROPOSED METHODOLOGY

The Proposed model has been implemented for the UW Madison GI Tract dataset with MRI scans of cancer patients. These scans have been carefully annotated to identify the small bowel, large bowel, and stomach regions. The proposed model for segmenting GI tract organs utilizes the PSPNet architecture, which consists of three main modules: the Feature Encoding Network, Pyramid Pooling Module, and Decoder, as shown in Figure 1. The feature encoding network consists of three main blocks: the Convolution Block, which extracts local patterns from the images. The Special Block incorporates spatial attention mechanisms, and the Average Pooling and Dense Layer Block refine the extracted features. To improve the model's performance, five transfer learning models, namely ResNext 50, Timm_Gernet_S, ResNet 34, EfficientNet B1, and MobileNet V2, were used as encoders for the PSPNet. The Pyramid Pooling Module is designed to gather context information at several scales, enabling a more comprehensive understanding of the image. On the other hand, the decoder utilizes upsampling techniques to build the final segmentation map that aligns with the original image resolution. This technology's primary objective is to precisely segment gastrointestinal (GI) tract organs, improving the interpretation and diagnosis of medical images.

IV. INPUT DATASET

The Institution of Wisconsin-Madison in Madison, Wisconsin, has made an MRI scan dataset available. This database contains 85 patients having scans of 1 to 6 days. Every day's scan has 144 or 80 images for diverse patients. The images in the dataset are in RLE encoding form. By using these RLE-encoded images, masks are produced using deep learning algorithms. The size of the images in this database is 224 × 224 × 3. The collection has 38,496 images in total. The images are divided into train and test with a ratio of 80:20, respectively. The numbers of train and test images are 30796 and 7700, respectively. Figures 2 (a), (b) & (c) show some sample images from the dataset, and figures 2 (d), (e), and (f) show their respective ground truth masks. Here, red indicates the large bowel, green indicates the small bowel, and blue indicates the stomach.

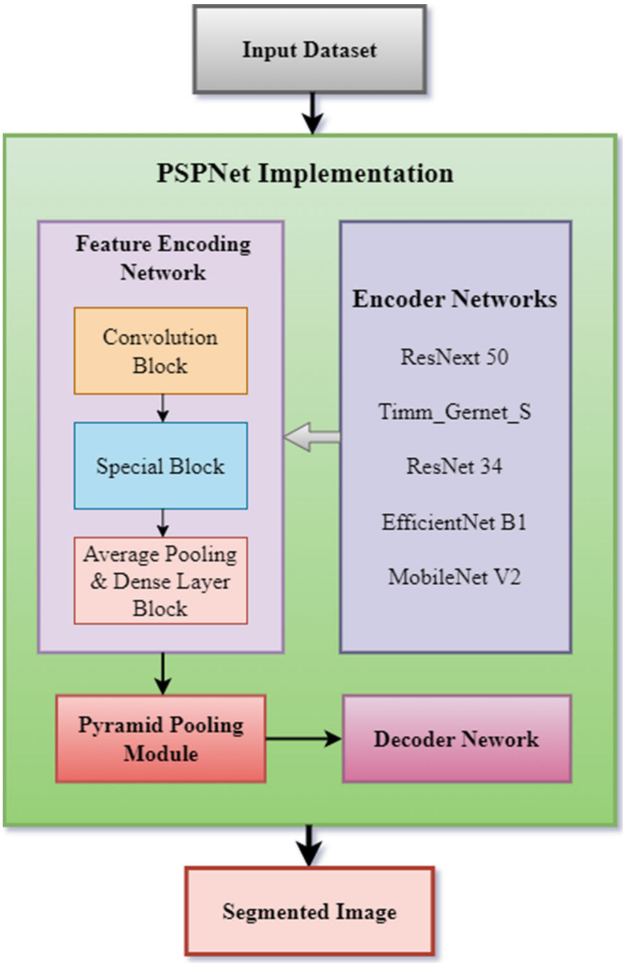


FIGURE 1. Proposed research methodology.

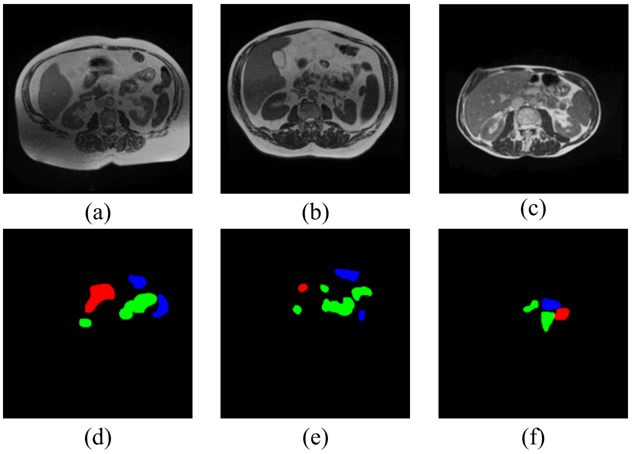


FIGURE 2. UW Madison GI tract dataset. (a), (b) & (c) Input images &(d), (e) & (f) Ground truth masks.

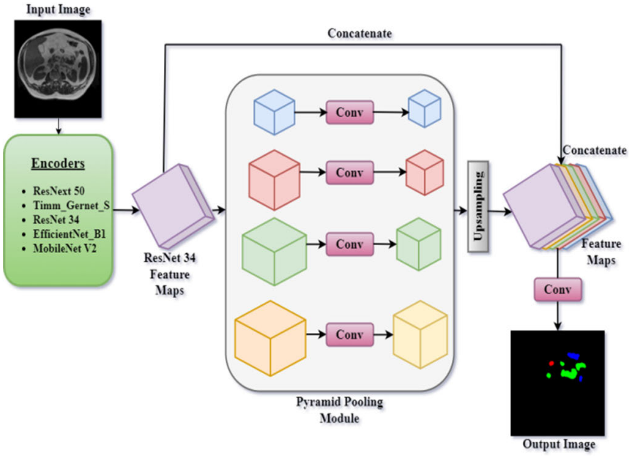


FIGURE 3. PSPNet architecture.

V. PSPNet IMPLEMENTATION

Here, Pyramid Scene Parsing Network (PSPNet) is implemented to segment GI tract from UW Madison GI Tract dataset images. PSPNet is a semantic image segmentation convolutional neural network architecture. PSPNet is intended to identify every pixel in an image with its matching object class, a fundamental problem in computer vision. Its basic concept is to employ a pyramid pooling module that gathers contextual information at multiple scales to increase segmentation accuracy. The pyramid pooling module comprises many parallel layers with varying kernel sizes concatenated and fed into a CNN for classification [27]. Figure 3 depicts the PSPNet model's architecture.

PSPNet can be wisely used for biomedical image segmentation [28], [29], [30]. The PSPNet (Pyramid Scene Parsing Network) architecture comprises a pyramid pooling unit and the characteristic mixture element. The pyramid pooling unit is a critical PSPNet component that gathers various sizes of contextual information. It is made up of numerous parallel pooling layers with varying kernel sizes that are used to extract features at different scales. Each pooling layer generates a feature map with another resolution. These feature maps are then concatenated and sent through a convolutional layer to form a single feature map with a broad receptive field. The feature fusion module combines the pyramid pooling module's high-level features and the lower-level features from the network's earlier layers. This aids in the preservation of spatial information and improves segmentation accuracy. The feature fusion module comprises convolutional layers and skips connections that concatenate the feature maps from the module with the outputs from the network's prior layers. PSPNet's architecture is based on a Fully Convolutional Network (FCN) with a deep encoder and a decoder. The encoder comprises multiple convolutional layers that extract features from the input image.

The architecture mainly consists of three components: a feature encoding network, a pyramid pooling module, and a decoder network. The working and description of each element are explained below.

A. FEATURE ENCODING NETWORK

The input image is fed to the feature encoding network to get the feature map. The feature encoding network consists of several convolutional blocks with residual connections.

The output of the feature encoding network is upsampled to the original image size using bilinear interpolation. The upsampled feature map is concatenated with the pyramid pooling module's matching feature maps. The concatenated feature maps run the final segmentation mask through many convolutional blocks with residual connections to get the last segmentation mask.

Here, five transfer learning models named ResNext 50 [31], Timm_Gernet_S [32], ResNet 34 [33], EfficientNet B1 [34], and MobileNet V2 [35] were used as the encoder of the PSPNet for getting the feature map F(X).

Let X be the input image with W x H channels, and Y be the output feature map with W' x H' channels and K channels, where K is the number of filters in the final convolutional layer. Three blocks process the input image X: convolution, special, global average pooling, and dense layer.

- ***CONVOLUTION BLOCK***

First, a convolutional layer with F filters, a kernel size of K x K, a stride of S, and P padding is applied to the input X. The output is then subjected to a batch normalization layer and a rectified linear unit (ReLU) activation function. Let Z1 serve as this block's output. The encoder type will determine the values of F, K, S, and P. The PSPNet was encoded using the ResNext 50, Timm Gernet S, ResNet 34, EfficientNet B1, and MobileNet V2 transfer learning models to get the feature map. Table 1 provides the parameter values utilized by various encoders.

- ***SPECIAL BLOCK***

Depending on the type of encoder used, it consists of a special convolution block. The block might be a residual block for ResNet 34 and ResNext 50, a gernet block for Timm_Gernet_S, an inverted mobile block for the Efficient Net B1, or a depth-wise separable convolution block for the MobileNet V2 encoder. Table 2 provides a thorough overview of these components.

- ***AVERAGE POOLING AND DENSE LAYER BLOCK***

To obtain a feature vector of size, the output of the last special block is sent through a global average pooling layer. A dropout layer with a rate of 0.5 is added to the feature vector to prevent overfitting. A dense layer with a softmax activation is used in the feature vector to achieve the final output.

B. PYRAMID POOLING MODULE

The input X is first passed through a feature encoding network whose output is fed into a pyramid pooling module, which extracts features at multiple scales. Let P_k(X) be the output of the k-th pooling layer, where k = 1, 2, 3, 6. For each pooling layer, the feature map is first divided into non-overlapping regions of size (W_k, H_k), where

$$W\_k = \mathrm{ceil}(W/k) \qquad (1)$$

$$H\_k = \mathrm{ceil}(H/k) \qquad (2)$$

TABLE 1. Parameter values of different encoders.

Name of Encoder	No. of Filters (F)	Kernel Size (K)	Stride (S)	Padding (P)	Extra layers Name
ResNext 50 [31]	64	7x7	2	-	Max pooling layer
Timm_Gernet_S [32]	16	3x3	2	-	Batch normalization and ReLU activation
ResNet 34 [33]	64	7x7	2	3	Batch normalization and ReLU activation
EfficientNet B1 [34]	32	3x3	2	-	Batch normalization and swish activation
MobileNet V2 [35]	32	3x3	2	-	Batch normalization and ReLU activation

Here, W and H represent the width and height of the input feature map, and k is the pooling scale factor. By dividing the original dimensions by k and taking the ceiling of the result, these equations ensure that pooling operations capture contextual information at various spatial resolutions. In practice, the input feature map is pooled at multiple scales (commonly 1×1, 2×2, 3×3, and 6×6), resulting in feature maps with dimensions W_k and H_k. These pooled feature maps are then resized back to the original dimensions, forming a pyramid of feature maps that collectively capture multi-scale context, which is essential for accurate image segmentation in the PSPNet model.

Then, a global average pooling operation is applied to each region to obtain a feature vector of dimension D, where D is the number of feature channels. The feature vectors are concatenated along the channel dimension, resulting in a pooled feature map P(X) of size (W', H,' 4D).

C. DECODER NETWORK

In the previous sections, the feature encoding network output is upsampled using bilinear interpolation to the original image size. Let F(X) be the output of the feature encoding network, which is upsampled to get the original image size. Then, the upsampled feature map is concatenated with the corresponding feature maps from the pyramid pooling module (P_k(X)), where P_k(X) is the output of the k-th pooling layer and k = 1, 2, 3, 6. The outputs are passed through several convolutional blocks with residual networks to get the final segmentation mask Y.

Let D(P(F(X)) be the concatenation of the upsampled feature map P(F(X)) and the corresponding feature map from the pyramid pooling module P_k(X), where k = 1, 2, 3, 6. Then, the final segmentation mask Y is obtained as:

$$\begin{aligned} Y=&\mathrm{Conv}(D(P(F(X)),P\_1(X))+\mathrm{Conv}(D(P(F(X)), P\_2(X)) \\ &+\mathrm{Conv}(D(P(F(X)), P\_3(X))+\mathrm{Conv}(D(P(F(X)), \\ &P\_6(X))) \end{aligned} \qquad (3)$$

where Conv denotes a convolutional layer with appropriate parameters.

The capacity of PSPNet to gather contextual information at multiple scales is crucial for effectively recognizing the borders of different structures and tissues in biomedical image segmentation. In this proposed work, PSPNet (Pyramid Scene

TABLE 2. Detailed description of special blocks for different encoders.

Encoder Name	Special Block Name	Details
ResNext 50 [31]	Residual Block	It is divided into four phases, each containing many leftover blocks. Each residual block has two convolutional layers using a 3x3 filter size, a batch normalization layer, a ReLU activation function, and a skip connection that adds the input to the output of the second convolutional layer. The cardinality and breadth of the network dictate the number of filters in the convolutional layers.
Timm_gernet_s [32]	Gernet Block	It is made up of many grouped convolutions with 3x3 filters, which are then followed by batch normalization and ReLU activation functions. The number of groups in each block increases progressively from 1 to 8. The filters in each group are equal to the groups divided by the number of input channels.
ResNet 34 [33]	Residual Block	It is made up of two convolutional layers with 3x3 filters, which are followed by batch normalization and ReLU algorithms.
EfficientNet B1[34]	Mobile Inverted Block	It comprises numerous inverted bottleneck blocks, also known as mobile inverted residual blocks, intended to improve the network's efficiency and accuracy. Each inverted bottleneck block has a 1x1 convolutional layer, a 3x3 depthwise convolutional layer, and another 1x1 convolutional layer, all with varying numbers of filters and kernel sizes.
MobileNet V2[35]	Depthwise Separable Convolution Block	It comprises many depthwise separable convolutional blocks, each with a depthwise convolution with a kernel size of 3x3, a batch normalizing layer, a ReLU activation function, a pointwise convolution with 1x1 filters, a batch normalization layer, and a ReLU activation function. The width multiplier of the network determines the number of filters in the pointwise convolution.

Parsing Network) is used to segment gastrointestinal organs such as the stomach, small bowel, and large bowel in the gastrointestinal tract, which may assist the radio oncologist in speeding up cancer treatment.

VI. RESULTS AND DISCUSSION

The proposed model has been evaluated using five encoders: ResNext 50, Timm_gernet_s, ResNet 34, EfficientNet B1, and MobileNet V2. These encoders are transfer learning models that are trained on the imagenet dataset. The encoders were used for the downsampling in the PSPNet model to speed up the processing and increase results. The proposed model was implemented using a batch size of 128, and number of epochs used were 15. The encoders were taken from the smp library of PyTorch. All the simulations were performed using PyTorch and the Google Collab platform using Python. The following sections represent the results of implementing the PSPNet model with different encoders.

A. RESULTS WITH ResNext 50

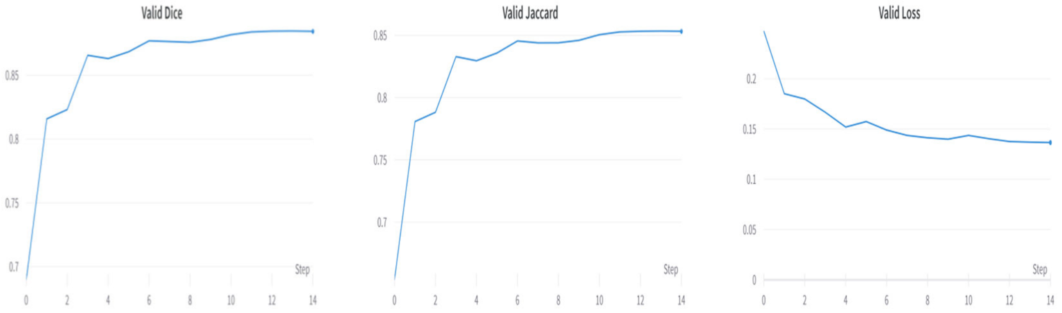
In Figure 4, a set of plots illustrates the training and validation performance of the GI tract organ segmentation model utilizing the ResNext 50 encoder. Figure 4(a), representing "Validation Dice," showcases a smoothly progressing curve, indicating consistent improvement in segmentation accuracy. Similarly, in Figure 4(b), "Validation Jaccard" also presents a smooth curve, suggesting a stable and high level of overlap between the model's predictions and the ground truth masks. Figure 4(c), depicting "Validation Loss," demonstrates a decreasing trend, with the curve consistently lowering, signifying practical model training. Importantly, these curves collectively indicate that the ResNext 50 encoder contributes to a highly performing model, as evidenced by achieving the highest values for Dice and Jaccard coefficients and the lowest value for validation loss, thus demonstrating the model's capacity to produce accurate and precise organ segmentations with minimal fluctuations.

B. RESULTS WITH Timm_Gernet_S

Figure 5 presents a comprehensive view of the training and validation process for the GI tract organ segmentation model using the Timm_Gernet_S encoder. In Figure 5(a), the "Validation Dice" curve illustrates the model's Segmentation accuracy, measured by the Dice coefficient, shows fluctuations in performance. Figure 5(b) depicts the "Validation Jaccard" curve, which assesses the overlap between predicted and ground truth masks with fluctuations indicating variations in segmentation quality. Finally, Figure 5(c) showcases the "Validation Loss" curve, which measures the overall error during validation, offering insights into how well the model generalizes to unseen data. These fluctuations across all three curves suggest that the model's performance fluctuates during training and validation, indicating potential challenges in achieving consistent and precise organ segmentation, which factors like dataset complexity, hyperparameters, or image characteristics could influence.

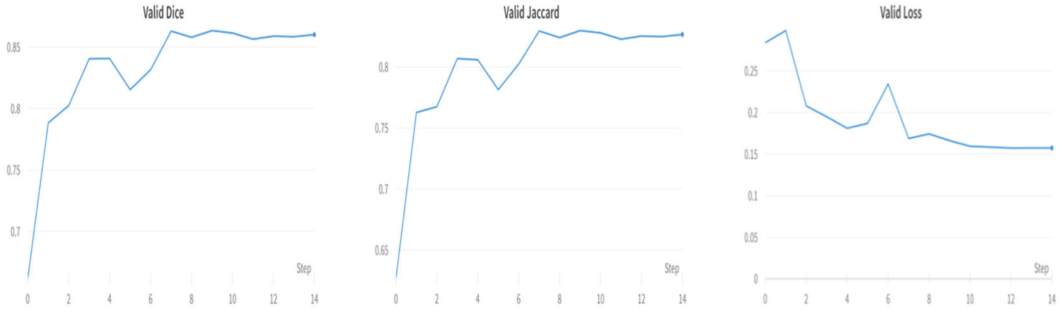
C. RESULTS WITH ResNet 34

Figure 6 presents a set of curves to illustrate the training and validation performance of the GI tract organ segmentation model using the ResNet 34 encoder. Figure 6(a), labeled "Validation Dice," portrays the model's segmentation accuracy, measured by the Dice coefficient. Despite some fluctuations, the curve indicates that the model consistently achieves excellent results for validation, suggesting strong agreement between its predictions and the actual organ boundaries. In Figure 6(b), "Validation Jaccard" displays a metric assessing the overlap between predicted and ground



(a) (b) (c)

FIGURE 4. Plots for ResNext 50 model (a) Validation dice, (b) Validation jaccard, and (c) Validation loss.



(a) (b) (c)

FIGURE 5. Plots for Timm_Gernet_S model (a) Validation dice, (b) Validation jaccard, and (c) Validation loss.

truth masks, with fluctuations but overall firm performance. Figure 6(c), representing "Validation Loss," tracks the overall error during validation, and while it may show some fluctuations, the model maintains a high level of performance throughout. These observations collectively indicate that the ResNet 34 encoder contributes to a robust and accurate organ segmentation model with consistently strong validation results.

D. RESULTS WITH EfficientNet B1

Figure 7 illustrates the training and validation performance using the EfficientNet B1 encoder in a specific experimental setup. In particular, Figure 7(a) portrays the validation dice coefficient, Figure 7(b) displays the validation Jaccard index, and Figure 7(c) represents the validation loss. The validation dice coefficient and Jaccard index typically increase till epoch four; after that, it has achieved constant values, which shows that the model has converged to an optimized state.

The validation loss, shown in Figure 7(c), typically decreases during training, but its behavior may also exhibit a plateau, indicating that the model has converged to a reasonably optimized state. This information is crucial for understanding the model's training dynamics and can guide decisions regarding when to halt training to prevent overfitting.

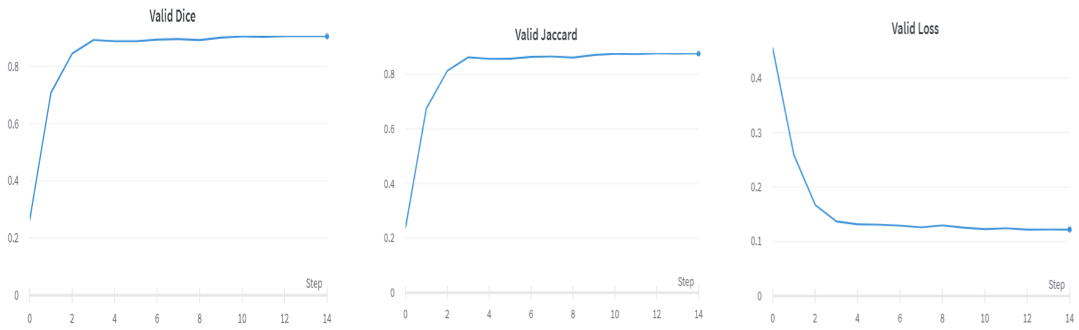
E. RESULTS WITH MobileNet V2

Figure 8 shows the training and validation processes employing the MobileNet V2 encoder. Specifically, in Figure 8(a), the validation dice coefficient is depicted, while Figure 8(b) showcases the validation Jaccard index, and Figure 8(c) represents the validation loss. What is notable from these plots is the substantial and oscillatory nature of the curves. The metrics in these figures exhibit frequent fluctuations, with values rising and falling.

Repeatedly throughout the training process. This pattern suggests that the model's performance is not consistently improving or converging to a stable solution. The oscillations may indicate challenges in training stability, possibly due to various factors such as learning rate adjustments, model architecture, or data quality issues. Addressing these fluctuations and achieving a more stable and steadily improving training trajectory may require further experimentation and optimization techniques to enhance the MobileNet V2-based model's performance.

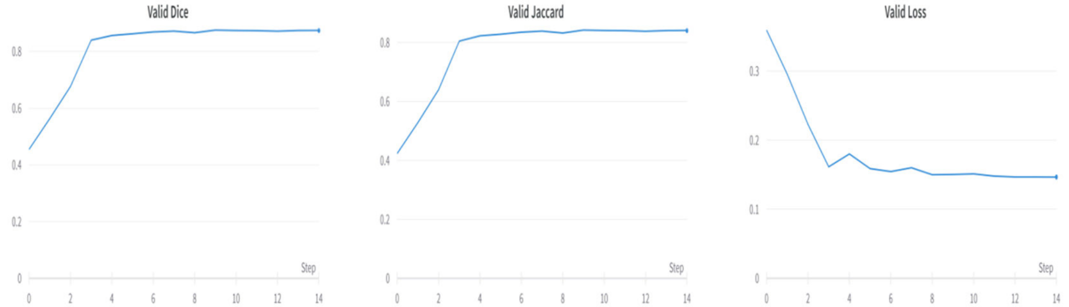
F. COMPARISON WITH DIFFERENT ENCODERS

This section presents a comprehensive comparison of various encoders within the PSPNet model through Figure 9, which includes four key aspects of model evaluation. In Figure 9(a) and Figure 9(b), the validation dice and Jaccard metrics are showcased, revealing that the ResNext 50 encoder achieves



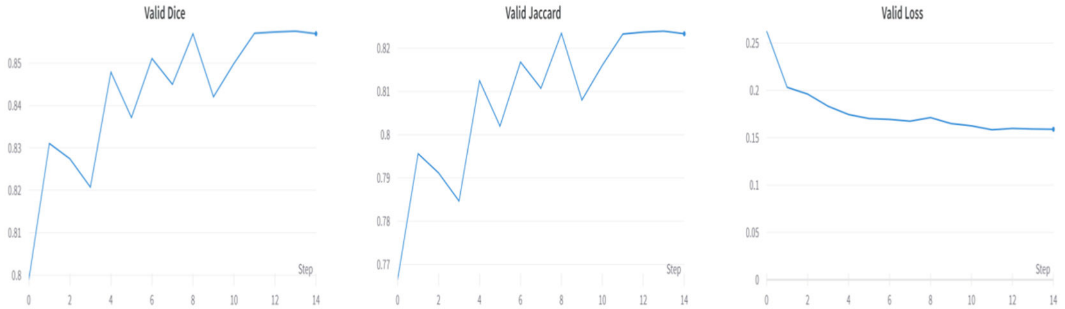
(a) (b) (c)

FIGURE 6. **Plots for ResNet 34 model (a) Validation dice, (b) Validation jaccard, and (c) Validation loss.**



(a) (b) (c)

FIGURE 7. **Plots for EfficientNet B1 model (a) Validation dice, (b) Validation jaccard, and (c) Validation loss.**



(a) (b) (c)

FIGURE 8. **Plots for MobileNet V2 model (a) Validation dice, (b) Validation jaccard, and (c) Validation loss.**

the highest performance. In contrast, the Timm_Gernet_S encoder performs less favorably, with the lowest values for both metrics. On the other hand, Figure 9(c) demonstrates that the ResNext 50 encoder results in the lowest validation loss, indicating better convergence and model fit. In contrast, the MobileNet V2 encoder leads to the highest loss values.

Regarding processing efficiency, as shown in Figure 9(d), the ResNext 50 model consumes the most time, likely due to its greater complexity and computational demands. At the same time, the MobileNet V2 encoder is the most computationally efficient, with the shortest processing time. These findings provide valuable insights into the trade-offs between different encoder choices within the PSPNet model, offering guidance for selecting an encoder based on specific priorities, whether accuracy, computational efficiency, or other factors, depending on the application's requirements. The comparison concludes that the ResNet 34 outperforms all the models regarding validation dice, Jaccard, loss, and processing time.

Table 3 comprehensively compares different encoders in terms of various performance parameters within the studied context. It is evident from the table that ResNext 50 stands out as the top performer, achieving impressive values of

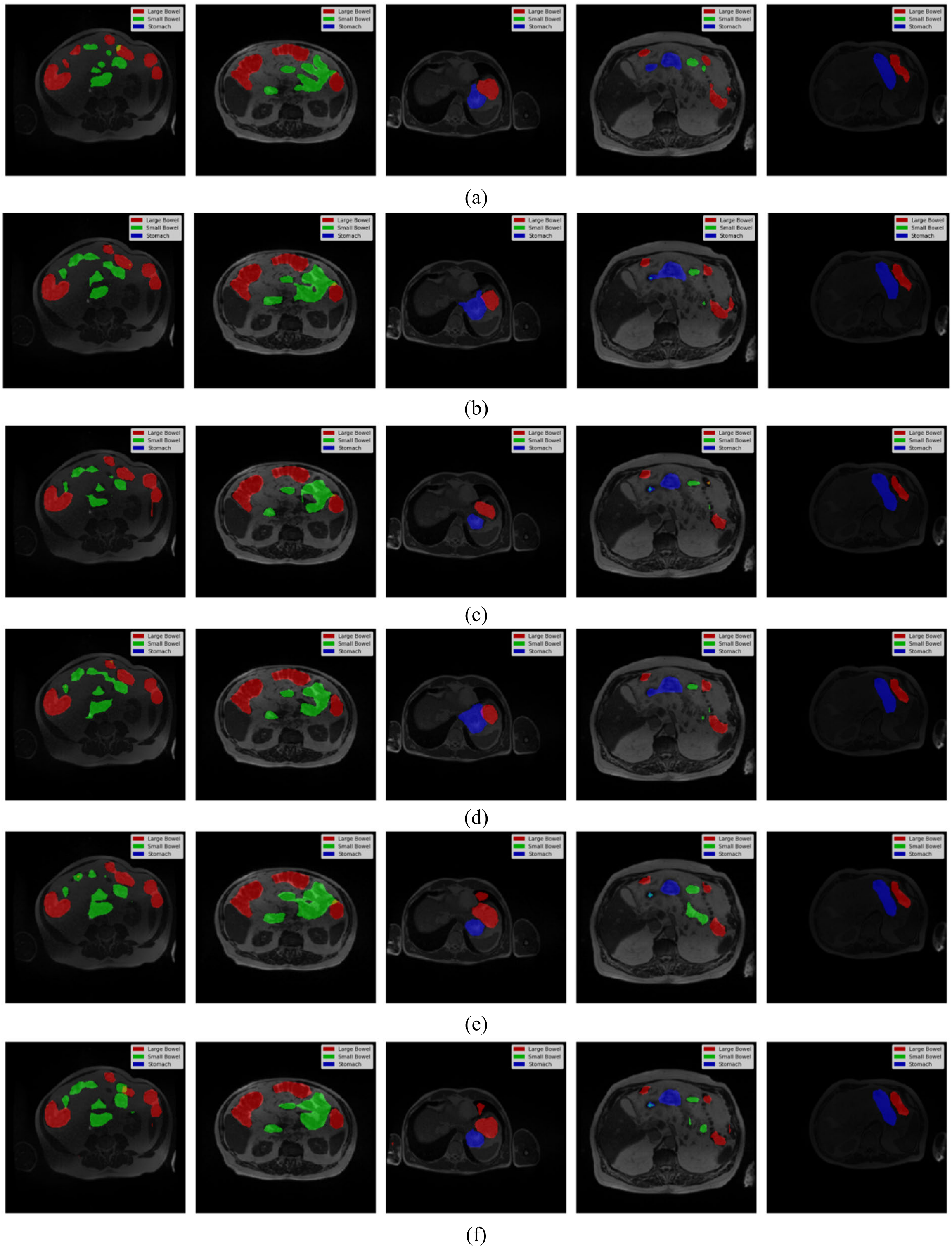


FIGURE 9. Visualization of segmented images with different feature encoding network (a) Ground truth masks, (b) Images predicted by ResNext 50, (c) Images predicted by Timm_gernet_s, (d) Images predicted by ResNet 34, (e) Images predicted by EfficientNet B1, and (f) Images predicted by MobileNet V2. Here, the red color shows the large bowel, the green color indicates the small bowel and the blue shows the stomach.

0.8931 for validation dice, 0.8596 for validation Jaccard, and 0.1294 for validation loss. However, it's worth noting that this exceptional performance comes at a cost in processing time, with a substantial duration of 5 hours and 35 minutes. Following closely behind, the ResNet 34 encoder demonstrates strong performance with values of 0.8842 for validation dice,

TABLE 3. Comparison of different encoders in terms of performance parameters.

Encoder	Valid Dice	Valid Jaccard	Valid Loss	Processing Time
ResNext 50	0.8931	0.8596	0.1294	5h 35m
Timm_gernet_s	0.8601	0.8265	0.1574	2h 24 m
ResNet34	**0.8842**	**0.8531**	**0.1365**	**2h 35m**
EfficientNetB1	0.8751	0.8423	0.1465	2h 28m
Mobilenet V2	0.8569	0.8233	0.1588	2h 22m

0.8531 for validation Jaccard, 0.0932 for train loss, and 0.1365 for validation loss. Notably, ResNet 34 achieves this commendable performance while maintaining a considerably shorter processing time of 2 hours and 35 minutes. This comparison highlights the trade-off between model performance and computational efficiency, providing valuable insights for selecting an appropriate encoder.

The PSPNet segmentation model exhibits limitations in terms of computational complexity, memory usage, and data requirements. To address these limitations, future research could focus on optimizing the computational efficiency of PSPNet, developing strategies to reduce memory, and exploring ways to mitigate data scarcity issues.

G. VISUALIZATION OF SEGMENTED IMAGES

Figure 9 provides a visual representation of the segmented images obtained from different encoders in the context of the UW Madison GI tract database scans. Figure 9(a), Figures 9(b), (c), (d), (e), and (f) depict the predicted masks generated by various encoders named ResNext 50, Timm_Gernet_S, ResNet 34, EfficientNet B1, and MobileNet V2. Ground truth mask serves as a reference, displaying the actual segmentation of the gastrointestinal tract with red color representing the large bowel, green indicating the small bowel, and blue denoting the stomach. Notably, all encoders exhibit good performance, effectively segmenting the masks predicted using ResNext 50 and ResNet 34 encoder, shown in Figure 9(b) and 9(d), respectively, closely resemble the ground truth masks, indicating high Jaccard index and dice coefficient in capturing the anatomical structures. The ResNext 50 takes more processing time as 5 hours 35 minutes, and ResNet 34 takes less time 2 hours 35 minutes, to process the model. This visual comparison underscores the remarkable performance of the ResNet 34 encoder in this particular task, providing valuable insights for medical image analysis and diagnosis.

VII. STATE OF ART COMPARISON

Table 4 compiles a comprehensive overview of image segmentation techniques employed on the UW Madison GI Tract dataset in 2022, accompanied by their respective outcomes. These techniques encompass diverse approaches applied to the medical imaging dataset, intending to delineate structures within the gastrointestinal (GI) tract accurately.

TABLE 4. State of art comparison for UW Madison GI tract dataset segmentation.

Ref. No./ Year	Techniques	Results
[19]/ 2022	UNet with Efficient Net B3	Jaccard-0.84
[20]/ 2022	SIA UNet	Jaccard- 0.83
[21]/ 2022	CNN Transformer	Dice- 0.79, Jaccard- 0.72
[22]/ 2022	UNet + Mask R-CNN	Dice- 0.51
[23]/ 2022	Residual Network	Jaccard – 0.75
[24]/ 2022	UNet on 2.5D	Dice- 0.36 Jaccard - 0.12
Proposed Model	**PSPNet + ResNet 34**	**Dice- 0.8842, Jaccard- 0.8531**

The methods featured include UNet with EfficientNet B3, which yielded an exceptionally high Jaccard score of 0.84, indicating its proficiency in segmentation.Another technique, SIA UNet, demonstrated a Jaccard score of 0.83, indicating its ability to perform segmentation satisfactorily. The CNN Transformer approach achieved a Dice coefficient of 0.79 and a Jaccard index of 0.72, suggesting its competence in delineating structures in the GI tract images. Conversely, the UNet + Mask R-CNN technique delivered a Dice score of 0.51, highlighting its performance at a moderate level. In contrast, the Residual Network produced an unusually low Jaccard score of 0.75, indicating room for improvement. Additionally, the UNet on 2.5D displayed a Dice coefficient of 0.36 and a Jaccard index of 0.12, suggesting that it may require enhancements for more accurate segmentation. The "Proposed Model," combining PSPNet and ResNet 34, demonstrated impressive results with a Dice coefficient of 0.8842 and a Jaccard index of 0.8531. Table 4 shows that the proposed model outperforms the state-of-the-art results for segmenting small bowel, large bowel, and stomach in the GI tract using the UW Madison GI tract dataset.

VIII. CONCLUSION

This study addresses the growing global challenge of rising gastrointestinal (GI) cancer cases by introducing a deep learning-based PSPNet model equipped with diverse pre-trained encoders tailored for the precise segmentation of healthy GI tract organs, encompassing the stomach, small intestine, and large intestine. Given the potential risks associated with radiation therapy in damaging healthy organs during GI cancer treatment, the primary objective of this model is to aid radiation oncologists in the swift and accurate delineation of these vital organs, ultimately enhancing the efficacy of therapy administration. Utilizing the UW Madison GI tract dataset, the outcomes underscore the model's effectiveness. The results reveal that the proposed PSPNet model with ResNet 34 as encoder outperforms the other feature encoding networks with validation dice as 0.8842, validation Jaccard as 0.8531, and validation loss as 0.1365. The model also takes the least time to implement, 2 hours and 35 minutes. Other encoders also demonstrate good performance, achieving a Dice value of 0.8931 for ResNext 50, 0.8601 for Timm_Gernet_S, 0.8751 for EfficientNet B1, and 0.8569 for MobileNet V2. Regarding Jaccard values, they attain 0.8596 for ResNext 50, 0.8265 for Timm_Gernet_S, 0.8423 for EfficientNet B1, and 0.8233 for MobileNet V2.

REFERENCES

[1] M. A. Khan, M. Rashid, M. Sharif, K. Javed, and T. Akram, "Classification of gastrointestinal diseases of stomach from WCE using improved saliency-based method and discriminant features selection," *Multimedia Tools Appl.*, vol. 78, no. 19, pp. 27743–27770, Oct. 2019, doi: 10.1007/s11042-019-07875-9.

[2] N. Sharma, A. Sharma, and S. Gupta, "A comprehensive review for classification and segmentation of gastro intestine tract," in *Proc. 6th Int. Conf. Electron., Commun. Aerosp. Technol.*, Dec. 2022, pp. 1493–1499.

[3] *Cancer.org*. Accessed: Oct. 4, 2023. [Online]. Available: https://www.cancer.org/cancer/stomach-cancer/about/key-statistics.html

[4] J. Grand, B. J. Richard, and M. F. Watkins, "Development of the human gastrointestinal tract: A review," *Gastroenterology*, vol. 70, no. 5, pp. 790–810, 1976.

[5] B. van Ginneken, C. M. Schaefer-Prokop, and M. Prokop, "Computer-aided diagnosis: How to move from the laboratory to the clinic," *Radiology*, vol. 261, no. 3, pp. 719–732, Dec. 2011.

[6] J. Sykes, "Reflections on the current status of commercial automated segmentation systems in clinical practice," *J. Med. Radiat. Sci.*, vol. 61, no. 3, pp. 131–134, Sep. 2014, doi: 10.1002/jmrs.65.

[7] M. Mittal, L. M. Goyal, S. Kaur, I. Kaur, A. Verma, and D. J. Hemanth, "Deep learning based enhanced tumour segmentation approach for MR brain images," *Appl. Soft Comput.*, vol. 78, pp. 346–354, May 2019.

[8] C. E. Prema, S. Suresh, M. N. Krishnan, and N. Leema, "A novel efficient video smoke detection algorithm using co-occurrence of local binary pattern variants," *Fire Technol.*, vol. 58, no. 5, pp. 3139–3165, Sep. 2022.

[9] S. Singh, A. K. Aggarwal, P. Ramesh, L. Nelson, P. Damodharan, and M. T. Pandian, "COVID 19: Identification of masked face using CNN architecture," in *Proc. 3rd Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Aug. 2022, pp. 1045–1051.

[10] S. F. Heavey, E. J. Roeland, A. M. P. Tipps, B. Datnow, and J. K. Sicklick, "Rapidly progressive subcutaneous metastases from gallbladder cancer: Insight into a rare presentation in gastrointestinal malignancies," *J. Gastrointest. Oncol.*, vol. 5, no. 4, pp. E58–E64, 2014, doi: 10.3978/j.issn.2078-6891.2014.023.

[11] B. Li and M. Q.-H. Meng, "Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 323–329, May 2012, doi: 10.1109/TITB.2012.2185807.

[12] M. Zhou, G. Bao, Y. Geng, B. Alkandari, and X. Li, "Polyp detection and radius measurement in small intestine using video capsule endoscopy," in *Proc. 7th Int. Conf. Biomed. Eng. Informat.*, Oct. 2014, pp. 237–241.

[13] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, and P. C. de Groen, "Polyp-alert: Near real-time feedback during colonoscopy," *Comput. Methods Programs Biomed.*, vol. 120, no. 3, pp. 164–179, Jul. 2015, doi: 10.1016/j.cmpb.2015.04.002.

[14] Q. Li, G. Yang, Z. Chen, B. Huang, L. Chen, D. Xu, X. Zhou, S. Zhong, H. Zhang, and T. Wang, "Colorectal polyp segmentation using a fully convolutional neural network," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017.

[15] Q. Nguyen and S.-W. Lee, "Colorectal segmentation using multiple encoder-decoder network in colonoscopy images," in *Proc. IEEE 1st Int. Conf. Artif. Intell. Knowl. Eng. (AIKE)*, Sep. 2018, pp. 208–211.

[16] W. Dijkstra, A. Sobiecki, J. Bernal, and A. Telea, "Towards a single solution for polyp detection, localization and segmentation in colonoscopy images," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (SCITEPRESS)*, 2019, pp. 616–625.

[17] N.-Q. Nguyen, D. M. Vo, and S.-W. Lee, "Contour-aware polyp segmentation in colonoscopy images using detailed upsampling encoder-decoder networks," *IEEE Access*, vol. 8, pp. 99495–99508, 2020, doi: 10.1109/ACCESS.2020.2995630.

[18] D. Jha, S. Ali, K. Emanuelsen, S. A. Hicks, V. Thambawita, E. Garcia-Ceja, M. A. Riegler, T. de Lange, P. T. Schmidt, H. D. Johansen, D. Johansen, and P. Halvorsen, "Kvasir-instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy," in *MultiMedia Modeling*. Cham, Switzerland: Springer, 2021, pp. 218–229.

[19] M. Sharma, "Automated GI tract segmentation using deep learning," 2022, *arXiv:2206.11048*.

[20] R. Ye, R. Wang, Y. Guo, and L. Chen, "SIA-unet: A unet with sequence information for gastrointestinal tract segmentation," in *Proc. Pacific Rim Int. Conf. Artif. Intell.* Cham, Switzerland: Springer, 2022, pp. 316–326.

[21] P. Nemani and S. Vollala, "Medical image segmentation using LeViT-UNet++: A case study on GI tract data," 2022, *arXiv:2209.07515*.

[22] A. Chou, W. Li, and E. Roman, "GI tract image segmentation with U-Net and mask R-CNN," Image Segmentation With U-Net Mask R-CNN, Stanford Univ., Stanford, CA, USA, Tech. Rep. cs231n-164, 2022.

[23] H. Niu and Y. Lin, "SER-UNet: A network for gastrointestinal image segmentation," in *Proc. 2nd Int. Conf. Control Intell. Robot.*, New York, NY, USA: ACM, Jun. 2022, pp. 227–230.

[24] H. Li and J. Liu, "Multi-view unet for automated GI tract segmentation," in *Proc. 5th Int. Conf. Pattern Recognit. Artif. Intell. (PRAI)*, Aug. 2022, pp. 1067–1072.

[25] B. Chia, H. Gu, and N. Lui, "Gastrointestinal tract segmentation using multi-task learning," Stanford Univ., Stanford, CA, USA, Tech. Rep. CS231n-75, 2022.

[26] M.-I. Georgescu, R. T. Ionescu, and A.-I. Miron, "Diversity-promoting ensemble for medical image segmentation," 2022, *arXiv:2210.12388*.

[27] X. Long, W. Zhang, and B. Zhao, "PSPNet-SLAM: A semantic SLAM detect dynamic object by pyramid scene parsing network," *IEEE Access*, vol. 8, pp. 214685–214695, 2020.

[28] X. Zhu, Z. Cheng, S. Wang, X. Chen, and G. Lu, "Coronary angiography image segmentation based on PSPNet," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105897.

[29] M. Zhang, X. Li, M. Xu, and Q. Li, "Automated semantic segmentation of red blood cells for sickle cell disease," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 11, pp. 3095–3102, Nov. 2020, doi: 10.1109/JBHI.2020.3000484.

[30] Z. Deng, K. Zhang, B. Su, and X. Pei, "Classification of breast cancer based on improved PSPNet," in *Proc. IEEE/ACIS 6th Int. Conf. Big Data, Cloud Comput., Data Sci. (BCD)*, Sep. 2021, pp. 86–90.

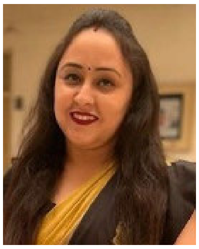
[31] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.

[32] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," 2020, *arXiv:2012.12877*.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[34] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.

[35] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.



NEHA SHARMA received the bachelor's degree in electronics and communication engineering from Kurukshetra University, Haryana, and the master's degree in electronics and communication engineering from Chitkara University, Punjab, where she is currently pursuing the Ph.D. degree in electronics and communication engineering. Her research interests include artificial intelligence, digital image processing, machine learning, deep learning, and computer vision.

SHEIFALI GUPTA is a Professor with the Chitkara University Research and Innovation Network (CURIN), Chitkara University, Punjab Campus, India. She specializes in the areas of digital image processing, pattern recognition, machine intelligence, biomedical image processing, agriculture based image processing, and deep learning. She has published more than 100 research articles and papers in reputed national and international journals and conferences. She has received the prestigious IRDP Award, in 2018, for remarkable achievements in teaching, research, and publications.

ADEL RAJAB received the bachelor's degree in computer science and information system and the master's and Ph.D. degrees in computer science and engineering from the University of South Carolina, USA. He is currently an Associate Professor with the College of Computer Science and Information System (CSIS); and the Vice-Dean of graduate studies for academic affairs with Najran University, Najran, Saudi Arabia. His research interests include robotics, drones, machine learning, and bioinformatic OBS networks.

MOHAMED A. ELMAGZOUB was born in Riyadh, Saudi Arabia, in 1985. He received the B.S. degree (Hons.) in electrical and electronic engineering from Omdurman Islamic University, Sudan, in 2007, and the master's degree in electrical communication engineering and the Ph.D. degree in electrical engineering (communications) from Universiti Teknologi Malaysia (UTM), Malaysia, in 2012 and 2016, respectively. He was a Telecommunication Engineer with the Communication Department, Dams Implementation Unit, Sudan, from 2008 to 2009. He is currently an Assistant Professor with the Department of Network and Communication Engineering, Najran University. He has authored ten ISI articles, two of them highly cited review articles in the area of passive optical networks and hybrid optical and wireless communication systems. His research interests include passive optical networks, hybrid optical and wireless communication systems, optical OFDM communications, and MIMO and next generation access networks.

KHAIRAN RAJAB received the bachelor's degree from the University of South Carolina, USA, the master's degree from the University of South Florida, USA, and the Ph.D. degree in computer science and engineering from the University of South Florida, under the supervision of Prof. L. A. Piegl. He is currently an Associate Professor, the Dean of e-learning and distance education, and the Head of the Network and Communication Engineering Department, College of Computer Science and Information Systems (CSIS), Najran University, Najran, Saudi Arabia. He has published over 16 research papers in high impact factor journals and reputable conferences. His research interests include geometric modeling, NURBS, data mining, network security, and cyber learning. He is a member of the College Council, CSIS. He is a reviewer and on editorial board of several research conferences and journals.

ASADULLAH SHAIKH (Senior Member, IEEE) received the B.Sc. degree in software development from the University of Huddersfield, England, U.K., the M.Sc. degree in software engineering and management from the University of Gothenburg, Sweden, and the Ph.D. degree in software engineering from the University of Southern Denmark. He is currently an Associate Professor, the Head of research, and the Coordinator of seminars and training with the College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia. He was a Researcher with UOC Barcelona, Spain. He has vast experience in teaching and research. He has more than 170 publications in the area of software engineering in international journals and conferences. His current research topics are UML model verification, UML class diagrams verification with OCL constraints for complex models, formal verification, and feedback technique for unsatisfiable UML/OCL class diagrams. He is an Editor of the *International Journal of Advanced Computer Systems and Software Engineering* (IJACSSE) and an international advisory board member of several conferences and journals.